

Construct(ion) and Context:
A response to methodological issues in studying character

Nancy L. Deutsch
National Academies of Science
Character Development Workshop
July 27, 2016

As I read Dr. Card's paper on methodological issues in studying character development, two words kept popping out at me: construct and context. Now, this might not be surprising. I am a qualitative and multiple methods inclined researcher. Studying how people construct meaning of and in their lives and how the social contexts in which they do so influence those meanings and, thereby, their behavior, are integral parts of what I study. That is not to say that quantitatively inclined researchers are not interested in these things. Rather, qualitative and quantitative researchers tend to approach constructs and contexts in different ways that make them more or less central, or at least central in different ways, to the questions that we ask about the world. And that is appropriate. Our questions drive our methods, and quantitative and qualitative methods have different, and complementary, strengths in terms of the types of knowledge and understanding that they address. Quantitative research is well suited for assessing the presence, amount, and prevalence of constructs related to character and to testing hypotheses about relationships between character and other constructs or actions. And Dr. Card's paper does a fine job at pointing out the methodological issues that arise when considering those features of character. Yet there are other elements of character development that are left unaddressed with such methods, as well as other questions that are raised. I think Dr. Card's paper hints at this when he discusses random effects models and how they acknowledge that some of the variability in statistical models is not statistical error, but is, as he says, "variability that is real." It is this "real" variability, and its implications for how we measure character development, on which I want to focus. And I will do so by digging into the terms construct and context and how I think they should inform our methods.

What struck me in this paper is how the implications of context for our methods differ widely depending on how we define the word "construct." Which brings me to why

the word construct struck me so powerfully. In developmental research, we continually refer to constructs. We measure them. We assess their prevalence and levels. We consider different conceptualizations of them, and the effects of those conceptualizations on what we conclude about the constructs. We study the relationship between constructs and test interventions that we think will have an effect on a construct, like character, that we think is important for human development. Yet, seldom in our day-to-day research or program lives do we deeply interrogate the basic assumptions we make about the constructs we measure. Seldom do we acknowledge that all *constructs* are *social constructions*. And so it is the nature of character as a social construction, the ways in which the very construction of character as a construct may differ across contexts, and the implications of these for how we study character, that I am going to try to address.

When we talk about construct validity or measurement reliability, we are acknowledging that the instruments we use to assess a construct like character are human constructions, and thus subject to error. We also recognize that there can be multiple measures that all claim to study character, and that do so with equivalent levels of validity and reliability, but that focus on different aspects of character, as Figure 1 from Dr. Card's paper illustrates. Yet when we examine measures in this way we are still assuming that something called character exists in the world as a singular, objective fact. The issue we grapple with in measurement, then, is not whether character exists, but how close we are able to get to the "real" or "true" (and I use quotation marks intentionally) construct of character. But all constructs in social science are human constructions and, as such, are bound by our social, historical, and political times and lenses.

Take psychopathology as an example. Up until 1973, homosexuality was defined as a sociopathic personality disorder in the DSM. Thus, the construct of homosexuality included mental illness as one of its core defining features. Over the past forty years, the construct of homosexuality has evolved. In 2009, the American Psychological Association released a resolution making clear that homosexuality is a "normal and positive variation of human sexuality" and opposing the use of conversion therapies. And today the construct of homosexuality is still being redefined. The very terms we use today, LGBTQQ, reflect the still broadening definition of that early construct in accordance with people's lived realities. So constructs change.

Yet how we define, and therefore measure, a construct at any given point in time has implications for who we consider “high” or “low” in that domain, and therefore how and with whom we intervene to affect change in that domain. Certainly the change in stance towards conversion therapies is one example. Another, and one tied distinctly to the link between definition and measurement, is moral development. In the late 1950’s, Lawrence Kohlberg began studying moral development in samples of White men. From this research, he developed six stages of moral development through which he proposed that people progress, with each stage representing a higher level of moral reasoning. As women began to be included in the samples testing this theory, it was found that women tended to get “stuck” at stage three, a stage that focuses on interpersonal relationships, as compared to the “higher” levels that focus on social contracts and universal ethical principles. Now if we stopped there, we might just conclude that women are less moral than men and in need of intervention to reach higher levels of moral reasoning. I am not suggesting that this is what Kohlberg concluded, but merely that when we test measures on new populations and find differences, the danger of interpreting those differences as deficits exists. Carol Gilligan saw the differences as a definitional and methodological issue, not as a reflection of true deficits in human development. She noted that by developing the initial theory based on male samples, Kohlberg had inadvertently omitted the experiences of women, which were different from that of men, and which led to different, not necessarily deficit, emphases when considering moral dilemmas. What she was saying, is that the context of women’s lives matter. So she went out and studied women, and developed the idea of an “ethos of care” as part of moral development. There has been a large literature on moral development since that time, and critiques of both Kohlberg and Gilligan’s work, including questions about whether there are, in fact, differences between men and women in this domain. Yet there have also been calls for including the concept of an “ethos of care” more directly within the concept of moral development as a result. And I believe an important lesson remains: that researchers and practitioners alike need to be careful about how we define “normative” in relation to any construct, as all constructs are, at least in part, products of the worldviews of the researchers who develop them and the populations with which the researchers develop their measures.

Now we might argue that this is indeed the whole point of testing the validity and reliability of measures in different populations. But in measurement development, we tend to look for measures that perform equivalently across populations or that have been used with populations that are similar to those whom we are studying. But we don't always consider whether a given population's lower performance on a measure that is otherwise valid and reliable is a meaningful indicator of deficit or, instead, is an indicator that the construct itself is different in that context. This is particularly important because we all view the world from the bodies in which we inhabit. Kohlberg experienced the world as a White male, as did the subjects from whose data he first developed his theory. Gilligan experienced the world as a White female, as did the subjects from whose data she developed her counter-theory. And anyone listening carefully will note that in both cases, there were still a lot people left out. While certainly non-White populations have been included in samples studying moral development since the theories were developed, their absence in the initial development means that what is defined as normative is still based on the experiences of White men and women, to say nothing of sample biases regarding social class, geography, sexuality, and other factors.

I think you can probably see where I am going with this. When we develop our measures founded on definitions of constructs based in the experiences of a narrow sub-set of people, we risk defining the world based on the experiences and expectations of that sub-set. The natural selection of programs, discussed by participants at this meeting, is part of this. Which ideas, theories, or programs survive depends in part on who has the power to make their ideas heard and to get money to fund their ideas. At a recent convening of researchers writing chapters for a volume about social-emotional development, a thought-provoking discussion occurred in response to a paper by Anne Gregory and Edward Fergus focused on racial disparities in school disciplinary practices and its links to social emotional learning. In the paper Gregory and Fergus suggest that a greater focus on teacher self-awareness is a necessary component of efforts to reduce the discipline gap between White and Black students. Further, they argue that "color blind" approaches to social emotional learning can inadvertently lead to understandings of SEL competence that are themselves blind to diverse expressions of SEL from students from marginalized backgrounds. To create a more equity-oriented conception of SEL, the authors argue, the

constructs underlying SEL should include factors such as an understanding of one's own conscious and unconscious biases as part of self-awareness or an awareness of the role of power and privilege in the actions of others as a part of social awareness. Consider now how changing the definitions of those SEL constructs would lead to different measurements of those constructs which may then also lead to different populations as being identified as in need of intervention.

Embedded in these last examples is the importance of context. Why do people have different experiences? Because context matters. As Dr. Card rightly points out, inter-informant reliability is only a reliable source of information if we assume that the construct itself is stable across contexts. Yet much human behavior is not. Research on student-teacher relationships, for example, demonstrates that students actually do behave differently with different teachers. Thus, a lack of inter-rater reliability between teachers may illustrate not instrument instability but actual differences in student and teacher experiences. Measurement equivalence assumes that a construct *should* operate "the same way across contexts, time, and/or groups." But if we believe that context matters, if we take Bronfenbrenner's social ecological theory seriously, then might we entertain the possibility that this is a false assumption? This is different than asking whether a person retains a core set of traits across contexts, as raised earlier in this meeting. It is asking if there is an essential center to the construct itself. Or, if a core center exists, if it presents in different ways or means different things across contexts. Take selfishness, a value we may try to discourage in people, as an example. Kathryn Edin's work with low-income mothers demonstrates how different people, with different life experiences, may define selfishness differently. In her research, she found that women from lower income backgrounds viewed upper income women's decisions to put off having children until they accomplished other life goals, as "selfish and unnatural," despite middle class narratives painting the decision to have a child earlier in life, and outside of marriage, as selfish. So whereas the definition of selfish for both sets of women may be the same (putting self before others), what is *defined* as selfish behavior differs.

I am not arguing that we should not try to quantitatively measure constructs such as character, that we deem to be important to human development and to the social good. Rather, I am arguing that we recognize that any definition we create of a construct such as

character, and any instrument we develop to measure it, reflects a shared understanding of that construct, not a reified, objective, and static reality to which we are holding a mirror. And it is here that qualitative and mixed methods can compliment quantitative measurements and help us identify and understand the implications of some of the potential measurement issues that arise when attempting to study character.

Qualitative methods, alone or in combination with quantitative methods, help us understand the meanings that people make of their worlds. Qualitative methods are typically naturalistic and help us understand people's perceptions and lived experiences. Rather than focusing on measuring the quantity or prevalence of a construct or phenomenon, qualitative researchers focus on understanding its meaning as it exists in context. You can think about the strengths of qualitative social science as akin to qualitative analysis in chemistry. Chemical qualitative analysis is an analysis that aims to identify the components of a mixture. Likewise, qualitative social science is useful in understanding the components of a social setting, i.e., the context. So rather than having to omit outliers or accept some level of random error, qualitative research can help us understand the meaning of those outliers and perhaps the reasons for those variations.

Thus, qualitative methods allow us to dig into variation that may be masked, or identified but not explained, by quantitative-only work. For example, in a recent paper on the consideration of race and ethnicity in research on youth development programs, Joanna Williams and I argue that researchers need more nuanced attention to race and ethnicity in our work. In particular, our reliance on over-simplified racial and ethnic groupings, and on statistical comparisons between those groupings, likely masks important differences in individual experiences. For example, when we compare Black to Latino/a to White youth on character development, we may miss important differences within each category. Black, for example, may include youth who are descendants of enslaved Americans, first generation immigrants from the Caribbean, or second generation African immigrants, just to name a few. Similarly, the category Latino/a may lump together fourth generation Cuban youth with first generation Mexican youth with Honduran refugees. And White may include third generation Irish youth, second generation Jewish youth, and first generation Bosnian youth. Due to the requirements of statistical power, seldom can a study interrogate differences within the categories we tend to use to compare people. The over-emphasis on

between-group differences, we argue, ignores important within group differences that help us understand how race and ethnicity operate as contexts of development that influence youth outcomes in different ways. Going back to my earlier point, qualitative methods could then help address issues of measurement equivalence by examining *how* and *why* measures may operate differently both within and between groups as well as over time, thus fine tuning both our understanding and measurement of a construct. In fact, whereas qualitative methods are often talked about in developmental work as exploratory, or as useful for measurement development, there is increasing acceptance of the idea that in-depth, rigorous qualitative and mixed methods work can yield important understandings about developmental pathways, or intraindividual change over time, and answer “why” and “how” questions by illuminating the ways in which a phenomenon occurs in the natural world.

Anthropologists understand that we learn how to be part of a culture or society by observing and listening. Children learn by observing and listening, then analyzing, synthesizing, and applying the information they take in. So, if we want to learn how we become people of character, we, too, need to observe and listen, analyze and synthesize, which is what qualitative researchers do. Indeed, qualitative research in after-school settings, including that of Reed Larson, and others at this meeting, has helped illuminate *how* youth develop character in these settings. Observations of youth’s relationships with adults and with each other have yielded important insights on what actions and structures promote character and other positive outcomes. And interviews with staff and youth have provided information on their perceptions, beliefs, and experiences. All of this information contributes to our understanding of how to structure and support organizations seeking to foster the character that most of us would agree is already *in* youth, but needs opportunities to flourish and be displayed.

And this is where I will leave you. I hope in a place not of despair, existential or otherwise, but of hope. Hope that whereas character development is a broad, complex construct, we have a multitude of methods on which we can draw to understand it. When researchers and practitioner acknowledge that character development is a social construction and that context matters in its definition, exhibition, and measurement, when we apply critical consciousness to our work, we can approach its study with more realistic

expectations about what to expect from our research, and stronger tools to help us gather the information we need to inform interventions and policy.